REVIEW

Dynamic risk assessment in patients with differentiated thyroid cancer

Fabian Pitoia and Fernando Jerkovich
Division of Endocrinology – Hospital de Clínicas – University of Buenos Aires, Ciudad de Buenos Aires, Argentina
Correspondence should be addressed to F Pitoia: fpitoia@intramed.net

Abstract

The stratification of patients with differentiated thyroid cancer based on their initial risk of recurrence, according to specific clinical, histopathological and perioperative data, is an important starting point for tailoring the follow-up during the first 1–2 years after initial therapy (surgery with or without radioiodine ablation). However, risk of recurrence re-stratification based on new clinical data that becomes available after considering the response to treatment (dynamic risk assessment) provides a more accurate prediction of the status at final follow-up and a more individualized approach. In this review, we summarized the available data regarding dynamic risk of recurrence and the suggested management of differentiated thyroid cancer patients according to the response to treatment. The use of this strategy is crucial to avoid overtreatment and intensive follow-up of the vast majority of patients who will have a very good prognosis and, on the other hand, focus therapeutic efforts on those patients who will have a worse prognosis. In the future, molecular biology analysis of the tumors and well-designed prospective studies will probably refine the risk of recurrence prediction.

Introduction

Differentiated thyroid carcinoma (DTC) is the most common endocrine neoplasia, accounting for nearly 2% of all types of cancer (Kitahara & Sosa 2016). Papillary thyroid carcinoma is the most common variant of thyroid cancer, followed by the follicular variant (Kitahara & Sosa 2016). Multiple staging systems have been developed to predict the risk of mortality in patients with DTC. The 2015 American Thyroid Association (ATA) guidelines (Haugen et al. 2016) recommend the AJCC/UICC TNM staging system for DTC patients based on its utility in predicting disease mortality. The estimated 10-year disease-specific survival (DSS) rate considering the recently published TNM 8th edition is 98–100% for stage I, 85–95% for stage II, 60–70% for stage III and <50% for stage IV patients (Amin et al. 2017, Tuttle et al. 2017). Assuming that the great majority of DTC patients will have a 10-year overall survival above 95%, in the last decade, the European Thyroid Association (Pacini et al. 2006), the ATA (ATA Guidelines Taskforce et al. 2009) and the Latin American Thyroid Society (Pitoia et al. 2009) among other societies, adopted the risk of recurrence (RR) classification (Fig. 1). These stratification systems take into account the histopathological report, intraoperative findings, as well as preoperative and immediate postoperative relevant data. This information allows physicians to tailor the response to treatment and then, the need for a close follow-up, mainly based on the probability of having persistent/recurrent structural disease.

While this initial RR stratification settles a starting point to guide an initial follow-up course for DTC patients, it represents a static picture that does not consider the response to initial treatment after thyroid surgery with or
without radioiodine remnant ablation (RA) (Tuttle et al. 2010). This re-stratification of the initial RR based on new clinical data that becomes available during the initial follow-up is called dynamic risk assessment. As expected, this strategy would provide a more accurate prediction of the RR and a more individualized approach (Tuttle et al. 2010).

### Using initial risk assessment in differentiated thyroid cancer

After completing initial therapy (thyroid surgery with or without RA), patients with a diagnosis of DTC are classified as having low, intermediate or high RR (Fig. 1).

The 2009 ATA RR stratification system (Fig. 1) was validated in different cohorts of DTC patients around the world (Tuttle et al. 2010, Castagna et al. 2011, Vaisman et al. 2012, Pitoia et al. 2013). These studies demonstrated that the percentage of structural incomplete response (evidence of local or distant persistent/recurrent disease) at the end of follow-up was 3–9% for low-risk patients, 13–45% for intermediate-risk patients and 60–80% for high-risk patients. In addition, the percentage of excellent response (no evidence of clinical or biochemical disease at final follow-up) was 78–91% in low-risk patients, 52–64% in intermediate-risk patients and 31–32% in high-risk patients (Tuttle et al. 2010, Castagna et al. 2011, Vaisman et al. 2012, Pitoia et al. 2013, Haugen et al. 2016).

In 2016, this 2009 ATA risk stratification system was updated (Fig. 1) in the modified risk stratification system from the 2009 ATA guidelines, which considered a series of additional clinical-pathological situations: number of vascular invasion foci, number and size of the lymph nodes involved, presence of extranodal extension and \textit{BRAF/TERT} mutational status, when available (Haugen et al. 2016). The Modified Stratification System from ATA 2009 guidelines was also validated in several cohorts of DTC patients.
patients around the world (Pitoia et al. 2015, Momesso et al. 2016, Abelleira et al. 2017, Park et al. 2017). In an investigation performed in 63 patients with low and intermediate RR from our hospital, we concluded that additional histopathological variables proposed by this new stratification system allowed for a greater refinement in the prediction of the different responses to treatment (Pitoia et al. 2018).

As another important feature, some authors consider the measurement of serum thyroglobulin (Tg) at 6 weeks postoperatively as a data belonging to the initial RR assessment (Tarasova & Tuttle 2016). Although a cut-off point for this value has not been clearly established yet, a non-stimulated (under thyroid hormone therapy) Tg level greater than 5ng/mL would alert about the possibility of local or distant metastasis and could, for example, in low- and intermediate-risk patients, guide the decision for RA (Tarasova & Tuttle 2016).

A few studies have explored the age at diagnosis as a variable related to the RR in DTC patients. Our group analyzed the impact of age at the diagnosis of DTC on the frequency of structural incomplete response, considering the Modified Stratification System from ATA 2009 guidelines (Pitoia et al. 2017). In this study, we did not find any age threshold that could distinguish an older group of patients with a higher risk for structural incomplete response neither at the initial response to treatment nor at the end of follow-up. However, a tendency toward a lower frequency of structural incomplete response was observed at the end of follow-up in patients with intermediate and high RR less than 40 years, without statistical significance. Indeed, other authors demonstrated that age was a key predictor for the response to therapy in patients with high RR (Shah et al. 2018). Moreover, high-risk younger patients with a structural incomplete response to therapy had a significantly better disease specific survival (DSS) than older patients (74 vs 12%, respectively, P<=0.001) (Shah et al. 2018).

The RR classification system was also validated in patients with low and intermediate risk or recurrence who did not receive RA. With a median follow-up of 5–10 years, the percentage of structural disease in patients receiving total thyroidectomy without RA was less than 1–2% in low-risk patients and 8% in those with intermediate RR (Vaisman et al. 2011a, Durante et al. 2012, Schwartz et al. 2012, Momesso et al. 2016, Abelleira et al. 2017, Park et al. 2017, 2018, Cho et al. 2018), similarly to what it was reported in patients who received total thyroidectomy and RA.

This initial RR is an important starting point for the management of a patient during the first 1–2 years after initial therapy. However, additional data in the follow-up (Tg and anti-Tg values or structural disease detected by imaging) can change this static RR. For example, in a patient initially classified as having an intermediate RR who achieves an excellent response (absence of biochemical and structural disease) within 2 years after initial therapy, the current likelihood of having persistent disease will be lower than that expected with the initial static RR classification. This example shows that the real RR of a patient with DTC is a dynamic process that is modified according to the responses to therapy that the patient obtains during follow-up and that the absence of complete data to establish the initial RR could be overcome when we have the response to therapy (La Greca et al. 2017).

Dynamic risk assessment in differentiated thyroid cancer

Patients treated with total thyroidectomy and radioiodine RA

The dynamic RR approach implies the re-stratification of the initial RR of DTC patients considering the different responses to treatment: excellent, indeterminate, biochemical incomplete and structural incomplete (Table 1), using specific data obtained during follow-up: Tg and anti-Tg values, results of imaging studies, including neck ultrasound (US), RAI scans, computed tomography (CT) or 18-fluorodeoxyglucose positron emission tomography-CT (18-FDG PET/CT), and so forth, guided by initial RR assessment (Fig. 1).

Table 1  Response-to-therapy definitions according to the initial treatment performed in patients with differentiated thyroid cancer.

<table>
<thead>
<tr>
<th>Excellent</th>
<th>Total thyroidectomy + RA</th>
<th>Total thyroidectomy without RA*</th>
<th>Lobectomy*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-stimulated Tg level</td>
<td>Non-stimulated Tg level 0.2–5 ng/mL</td>
<td>Stable non-stimulated Tg level &lt;30 ng/mL</td>
</tr>
<tr>
<td></td>
<td>&lt;0.2 ng/mL</td>
<td>Undetectable anti-Tg</td>
<td>Undetectable anti-Tg</td>
</tr>
<tr>
<td></td>
<td>Stimulated Tg level</td>
<td>Stable or declining TgAb levels</td>
<td>Stable or declining TgAb levels</td>
</tr>
<tr>
<td></td>
<td>&lt;1 ng/mL</td>
<td>Nonspecific findings on imaging studies</td>
<td>Nonspecific findings on imaging studies</td>
</tr>
<tr>
<td></td>
<td>Undetectable anti-Tg</td>
<td></td>
<td>Negative imaging</td>
</tr>
<tr>
<td></td>
<td>Negative imaging</td>
<td></td>
<td>Negative imaging</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>Non-stimulated Tg level</td>
<td>Non-stimulated Tg level &gt;5 ng/mL</td>
<td>Non-stimulated Tg level &gt;30 ng/mL</td>
</tr>
<tr>
<td></td>
<td>0.2–1 ng/mL</td>
<td>Undetectable anti-Tg</td>
<td>Increasing anti-Tg levels</td>
</tr>
<tr>
<td></td>
<td>Stimulated Tg level</td>
<td>Stable or declining TgAb levels</td>
<td>Negative imaging</td>
</tr>
<tr>
<td></td>
<td>1–10 ng/mL</td>
<td>Nonspecific findings on imaging studies</td>
<td>Negative imaging</td>
</tr>
<tr>
<td></td>
<td>Stable or declining TgAb levels</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nonspecific findings on imaging studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biochemical incomplete</td>
<td>Non-stimulated Tg level</td>
<td>Non-stimulated Tg level &gt;5 ng/mL</td>
<td>Non-stimulated Tg level &gt;30 ng/mL</td>
</tr>
<tr>
<td></td>
<td>&gt;1 ng/mL</td>
<td>Increasing anti-Tg levels</td>
<td>Increasing Tg or anti-Tg levels</td>
</tr>
<tr>
<td></td>
<td>Stimulated Tg level</td>
<td>Negative imaging</td>
<td>Negative imaging</td>
</tr>
<tr>
<td></td>
<td>&gt;10 ng/mL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increasing anti-Tg levels</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Negative imaging</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structural incomplete</td>
<td>Structural or functional evidence of disease</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* TSH stimulation is not recommended in patients without radioiodine remnant ablation.

RA, radioiodine remnant ablation.

The dynamic risk approach was initially proposed by Michael Tuttle from the Memorial Hospital of New York (Tuttle et al. 2010). This was a retrospective study of 588 patients with DTC who received total thyroidectomy and RA, with a median follow-up of 7 years (range 1–15 years). They were initially stratified by the RR of the ATA 2009 guidelines (ATA Guidelines Taskforce et al. 2009). After this initial RR assessment, the best response to treatment obtained during the first 2 years after initial therapy was analyzed. After the re-stratification, among patients who had an initial excellent response, the probability of having structural incomplete response was reduced from 3 to 2% in low-risk patients, from 18 to 2% in intermediate-risk patients, and, which was more remarkable, from 66 to 14% in patients with high RR (Tuttle et al. 2010). On the contrary, having an initial structural or biochemical incomplete response increased the percentage of persistent structural disease at the end of follow-up to 13% in low-risk patients, 41% in intermediate-risk patients and 79% in high-risk patients (Tuttle et al. 2010). Moreover, the authors compared the static stratification considering the initial RR and the dynamic risk assessment using the proportion of the variance explained (PVE), which measures the ability of a classification system to predict the final response (structural incomplete vs excellent response). As expected, the PVE was 34% when the initial RR was considered and it increased to 84% when the responses to treatment during the first 2 years after initial therapy were analyzed.

These data were then validated in another retrospective cohort of 512 patients. In this study, the percentage of patients who achieved complete remission at the end of follow-up was 90% for those with low risk and 96% for those with low dynamic risk (patients who had an initial excellent response to therapy) (Castagna et al. 2011), and, a more important finding was that the percentage of complete remission in patients with intermediate-high-risk dropped from 60 to 27% in those with high dynamic risk (patients with initial biochemical persistence or structural incomplete response). The positive predictive value (PPV) was 72% for the dynamic risk stratification compared to 39% for the initial RR stratification, which means that the dynamic risk assessment had a better ability to predict the recurrence/persistence of disease. Accordingly, the recurrence rate after complete remission was low (2.7%), and there were no statistically significant differences between the initial low and high RR patients. This good prognosis in patients with an excellent initial response to treatment regardless of the initial RR was replicated in other studies (Hong et al. 2014, Pitoia et al. 2015, Kowalska et al. 2016, Jeon et al. 2018) (Table 2).

The dynamic risk stratification system also proved to be associated with the progression-free survival (PFS). In a retrospective cohort of 398 patients with DTC,
through a multivariate analysis, it was found that the factors involved in the dynamic risk (stimulated Tg values >1 ng/mL and TSH suppressed Tg values >1 ng/mL) but not those involved in static RR (extranodal extension, presence of lymphadenopathy and histological type), were associated with PFS (Haugen et al. 2014). More recently, a retrospective study of 762 patients with PTC measuring 1–4 cm revealed that dynamic risk assessment using data obtained from the first 2 years after initial treatment was an independent risk factor for structural recurrence during the follow-up period and that the disease-free survival (DFS) curves of each response-to-therapy category of the dynamic risk stratification system showed significant differences (<0.001) (Lee et al. 2018).

These results demonstrate the superiority of the dynamic risk in the prediction of the final response to treatment over the static initial RR. Evidence from observational and prospective studies validating the dynamic risk stratification system in patients with total thyroidectomy and RA is presented in Table 2.

Management implications of the response-to-therapy categories

As it happened with the initial RR, the response to treatment can be used to guide the management of patients with DTC. (Fig. 2).

An excellent response is generally achieved in 86–91% of patients with low RR, 57–63% of those with intermediate RR and in 14–16% of those with high RR who received RA (Tuttle et al. 2010, Castagna et al. 2011, Vaisman et al. 2012, Pitoia et al. 2013, Haugen et al. 2016, Kowalska et al. 2016, Lee et al. 2018, Schlumberger et al. 2018, Dehbi et al. 2019, van Velsen et al. 2019). When a low-risk patient has an excellent response to treatment, although it is reassuring, the therapeutic implication in the dynamic risk assessment does not change substantially, since the risk of structural incomplete response was initially low in this group of patients. From 631 patients who had a complete ablation in the ESTIMABL1 study, only one patient had a structural incomplete response during follow-up (0.16%) (Schlumberger et al. 2018). This is a very strong data coming from a prospective study, consistent with the previously published retrospective experience in patients who received RA.

In patients with intermediate RR, on the other hand, an excellent response implies a substantial reduction of the probability of a structural incomplete response compared to that expected from their initial RR: from 36–43% to 1–2% (Tuttle et al. 2010, Castagna et al. 2011, Vaisman et al. 2012, Pitoia et al. 2013, 2015). Another study (Trimboli et al. 2017) showed that the optimized threshold of baseline Tg (on T4-Tg) measured 6–12 months after initial treatment in patients with low

Table 2 Summary of studies analyzing dynamic risk assessment in patients with differentiated thyroid cancer treated with total thyroidectomy and radioiodine remnant ablation.

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of patients</th>
<th>Patients included</th>
<th>Median follow-up (months)</th>
<th>Persistent/recurrent disease based on each response-to-therapy category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuttle et al. 2010</td>
<td>588</td>
<td>All-risk patients</td>
<td>84</td>
<td>Exc 4% IR 0% BIR/SIR 57%</td>
</tr>
<tr>
<td>Castagna et al. 2011</td>
<td>512</td>
<td>All-risk patients</td>
<td>81.6</td>
<td>Exc 3.4% BIR/SIR 66%</td>
</tr>
<tr>
<td>Hong et al. 2014</td>
<td>398</td>
<td>All-risk patients</td>
<td>128</td>
<td>Exc 1.3% SIR 75.9%</td>
</tr>
<tr>
<td>Pitoia et al. 2015</td>
<td>149</td>
<td>Low- and intermediate-risk patients</td>
<td>72</td>
<td>Exc 1.6% BIR/SIR 31.8%</td>
</tr>
<tr>
<td>Kowalska et al. 2016</td>
<td>916</td>
<td>All-risk patients</td>
<td>84</td>
<td>Exc 1.2% IR/BIR/SIR 16.2%</td>
</tr>
<tr>
<td>Trimboli et al. 2017</td>
<td>201</td>
<td>Low- and intermediate-risk patients</td>
<td>29</td>
<td>onT4-Tg &lt;0.28 ng/mL: 0% onT4-Tg &gt;0.28 ng/mL: 52% Exc 0.5% IR 37.5% BIR 50% SIR 85.7%</td>
</tr>
<tr>
<td>Jeon et al. 2018</td>
<td>1359</td>
<td>Patients with excellent response</td>
<td>104.4</td>
<td>Exc 1%</td>
</tr>
<tr>
<td>Lee et al. 2018</td>
<td>667</td>
<td>PTC 1-4 cm</td>
<td>124</td>
<td>Exc 4.1 IR 17.6% BIR 53.4% SIR 81.5%</td>
</tr>
<tr>
<td>Schlumberger et al. 2018</td>
<td>726</td>
<td>Low-risk patients</td>
<td>64.8</td>
<td>Exc 0.16% All initial responses 0.5%</td>
</tr>
<tr>
<td>Dehbi et al. 2019</td>
<td>434</td>
<td>T1-T3</td>
<td>78.4</td>
<td>Exc 0%; all initial responses 4.8%</td>
</tr>
<tr>
<td>van Velsen et al. 2019</td>
<td>236</td>
<td>N0/N1/Nx M0 (TNM 6th ed.)</td>
<td>72</td>
<td>Exc 14%</td>
</tr>
<tr>
<td>Tian et al. 2019</td>
<td>767</td>
<td>High-risk patients</td>
<td>67.2</td>
<td>Pre-ablation s-Tg &lt;1 ng/mL and negative anti-Tg: 2.9%</td>
</tr>
</tbody>
</table>

*Thyroglobulin on levothyroxine treatment measured at 6–12 months after radioactive iodine; †Prospective trials.

BIR, initial biochemical incomplete response; Exc, initial excellent response; IR, initial indeterminate response; PTC, papillary thyroid carcinoma; SIR, initial structural incomplete response; s-Tg, stimulated thyroglobulin; Tg, thyroglobulin.
and intermediate RR was set by ROC curves analysis at 0.28 ng/mL. Having onT4-Tg <0.28 ng/mL at 6–12 months after treatment was associated with longer disease free survival and absence of relapse during follow-up (odds ratio 0.328, $P=0.0001$) Another study of 226 patients with low- and intermediate-risk PTC who had an excellent response found ultrasonographically suspicious neck lymph nodes in only 1.2% of patients at 3 years and in 1.8% at the last visit. Additionally, 75% of the nodal lesions were likely false positive and none required treatment (Grani et al. 2019). These data support the idea that low- and intermediate-risk PTC patients with an excellent response can be safely followed with clinical assessments and unstimulated serum Tg determinations with no need of routine neck ultrasonography assessment.

On the other side, in the few high-risk patients who achieve an excellent response, several authors have demonstrated that the percentage of structural incomplete response at the end of the follow-up would also be low (1–3%) (Castagna et al. 2011, Jeon et al. 2018). According to these investigations, an excellent response would be maintained at the end of follow-up, independently of the initial RR. On the contrary, other authors showed that an excellent response to treatment in patients with initial high RR supposes a RR of 5–15% (Tuttle et al. 2010, Shah et al. 2018, van Velsen et al. 2019). Therefore, a closer follow-up and additional structural or functional imaging would be necessary to rule out disease that may not be detected by US and Tg measurements prior to classify these patients as having an excellent response (Haugen et al. 2016, Tarasova & Tuttle 2016). Additionally, in a cohort of 67 ATA high-risk patients who achieved an immediate excellent response (pre-ablation stimulated Tg <1 ng/mL), which occurred in 9% of cases, the structural incomplete response was only 3% after a median follow-up of 5.6 years (regional lymph node metastases in all cases) (Tian et al. 2019). The suggested management for patients with an excellent response to treatment can be seen in Fig. 2.

The persistence of Tg levels >1ng/mL under levothyroxine therapy, stimulated Tg levels >10ng/mL or rising anti-Tg in the absence of structural disease (biochemical incomplete response) in patients who received RA occurs in 11–19% of low RR patients, 21–22% of those with intermediate RR and in 16–18% of those with

---

**Figure 2**
Risk of structural disease at the end of follow-up and management recommendations for each response-to-therapy category. *For some authors, the percentage of structural incomplete response at final follow up might be higher in patients with an initial high risk of recurrence and a more strict follow-up should be recommended in this group of patients (Tuttle et al. 2010, Shah et al. 2018). CT, computed tomography; 18-FDG-PET/CT, positron emission tomography with 18-FDG integrated with computed tomography; MKI, multikinase inhibitor therapy; US, ultrasound.
high RR (Tuttle et al. 2010, Castagna et al. 2011, Vaisman et al. 2012, Pitoia et al. 2013). In general, these patients have a good prognosis, given that the majority (56–68%) will achieve an excellent response at final follow-up, 19–27% will continue with detectable Tg or anti-Tg values and only 8–17% will develop structural disease during 5–10 years of follow-up (Tuttle et al. 2010, Castagna et al. 2011, Vaisman et al. 2011b, 2012, Pitoia et al. 2013, 2014). The management in this group of patients can also be observed in Fig. 2, but in general will depend on the trend of the Tg/anti-Tg levels over time. Rising Tg or anti-Tg levels should prompt the use for additional imaging methods (tomography, 18-FDG-PET/CT, for example). On the contrary, declining Tg or anti-Tg values would require only routine observation and neck US every 2–3 years. The TSH goal should be 0.5–1 mIU/L (Haugen et al. 2016).

The presence of a structural incomplete response is observed in only 2–6% of patients with low RR, in 19–28% of those with intermediate RR and in the majority (67–75%) of those with high RR (Tuttle et al. 2010, Castagna et al. 2011, Vaisman et al. 2012, Pitoia et al. 2013, Kowalska et al. 2016, Lee et al. 2018, Schlumberger et al. 2018, Debi et al. 2019). The majority of these patients (around two-thirds) will remain in this category despite additional therapy/ies (surgery, radioiodine). The site of the persistent disease, the rate of progression, the radioiodine avidity and the response to previous therapy will condition the patient’s management: surgery, external beam radiation therapy, tyrosine kinase inhibitors or, in selected cases, active surveillance. The target TSH value in this group is <0.1 mIU/L, except in cases of atrial fibrillation, where values between 0.1 and 0.5 mUI/L are acceptable (Fig. 2) (Haugen et al. 2016).

An indeterminate response to therapy (Table 1) occurs in 12–29% of patients with low RR, 8–23% in the intermediate RR group and only in 0–4% of those with high RR (Tuttle et al. 2010, Vaisman et al. 2012). A retrospective study of 128 PTC patients which investigated the presence of risk factors for indeterminate response showed that patients with a rhTSH-stimulated Tg greater than 2 ng/mL prior to radioiodine ablation were almost four times more likely to have an indeterminate response than those with a rhTSH-stimulated Tg ≤2 ng/mL (Jeong et al. 2019). The management usually consists of observing nonspecific findings or small elevations of the Tg or anti-Tg level until the patient can eventually be reclassified in any of the above categories, maintaining a TSH level in the 0.5–1 mIU/L range. Similar to what happens with the biochemical incomplete response, only 13–20% of cases will develop structural disease in 10 years of follow-up (Fig. 2) (Haugen et al. 2016). A recent study found that out of 145 DTC patients with indeterminate response-to-therapy followed up for a median time of 68 months, 9% had biochemical or structural incomplete response. In this subgroup of patients, unstimulated Tg values steadily increased (mean percentage change +27.1% at each follow-up visit) independently from the baseline levels (Malandrino et al. 2019). These data denote that in patients with indeterminate response, the trend of unstimulated Tg values may be useful to identify patients who will present with a higher risk for persistent disease.

**Patients treated with total thyroidectomy or lobectomy without radioiodine RA**

In the 2015 ATA guidelines, lobectomy is proposed as the initial surgical approach for tumors <1 cm without extrathyroidal extension and clinically N0, and may be sufficient for low-risk papillary and follicular carcinomas up to 4 cm in the absence of aggressive features (Haugen et al. 2016). Moreover, considering that RA is not routinely recommended in low RR patients and that it should be considered in those with intermediate RR, a significant percentage of patients with DTC will not receive RA after initial surgery.

The first validation of the dynamic risk assessment in patients with DTC without RA was performed in 2016 (Momesso et al. 2016), based on a retrospective cohort of 507 patients, 320 treated with total thyroidectomy and 187 with lobectomy. As expected, 85.4% of patients had low RR and 14.6% intermediate RR. The authors demonstrated that the percentage of persistence/recurrence of the disease at the end of the follow-up (median 8.4 years) was 0 in those with an excellent response, 1.3% in those who had an indeterminate response, 31.6% in patients with biochemical incomplete response and 100% in patients with structural incomplete response. In a similar cohort of non-ablated patients (n = 357, median follow-up 8.6 years) other group demonstrated that DFS was significantly different according to the initial response to treatment (Park et al. 2017). This clearly demonstrated that the dynamic risk stratification also allows to predict the risk of structural disease during follow-up in patients who did not receive RA.

It is worth to mention that although almost 18–30% of the patients who did not receive RA in the different cohorts have an initial indeterminate response, only 1.3–1.6% will have a final structural incomplete response.
Tuttle Demidchik Verburg). If we consider that

10.1530/ERC

In contrast, in

Park,

Similarly, a retrospective study

Table 1

Lee).

Interestingly, the postoperative serum Tg levels of 208 patients with low-risk PTC who underwent lobectomy was also evaluated, and it was found that there was no significant differences in serum Tg level changes between patients with or without recurrent disease (Park et al. 2018). Therefore, periodic measurements of serum Tg levels would not be useful for predicting recurrent disease in these patients and decreased or stable serum Tg levels do not guarantee a lower probability of recurrence. However, only two patients of the cohort from this study had Tg measurement >30ng/mL and both had recurrent disease, which probably means that this cut-off value would still be indicative of a high probability of persistent structural disease (Park et al. 2018). Similarly, a retrospective study of 619 patients with papillary thyroid cancer treated with lobectomy (median follow-up of 103 months), showed that the anti-Tg status, the Tg levels or the changing trend of Tg/anti-Tg levels were not significantly associated with structural recurrence (Cho et al. 2018). Moreover, in this study of patients treated with lobectomy alone, each response to therapy category was not well correlated with the incidence of structural recurrence, and the PVE of the dynamic stratification system in these patients was only 32.4%, which is similar to what it was reported for the initial RR assessment in patients who received total thyroidectomy and RA (Tuttle et al. 2010).

As expected, all patients with low and intermediate RR who did not receive RA and had a structural incomplete response after initial evaluation (1–2.9%) will maintain this response at the end of follow-up (Momesso et al. 2016, Abelleira et al. 2017, Park et al. 2017, Cho et al. 2018, Lee et al. 2018).

On the other hand, patients without RA who have an initial excellent response have very low percentages of structural incomplete response at the end of follow-up, similar to what happens in patients who receive RA. In fact, none of the patients from the cohort of Momesso et al. who had an initial excellent response presented with structural evidence of disease at final follow-up, and only 2/256 (0.78%) patients of the cohort from Park et al. who had an initial excellent response, had recurrence of the disease 4 and 5 years after lobectomy (Momesso et al. 2016, Park et al. 2017).

There is little evidence to guide TSH targets or the use of thyroid hormone in ATA low-risk patients who have undergone lobectomy. The 2015 ATA guidelines suggest a TSH goal of 0.5–2mU/L. However, a recent study demonstrated that postoperative serum TSH levels did not affect recurrence in low-risk PTC patients who underwent thyroid lobectomy (Lee et al. 2019). If we consider that 73% of patients will have TSH >2mU/L at 1 year after lobectomy (Cox et al. 2018), many patients following the ATA recommendation will be unnecessarily treated with levothyroxine. Therefore, the decision for TSH suppression in this group of patients should be made in terms of patient-specific considerations (Lee et al. 2019). A review on the indications and follow-up of patients undergoing lobectomy for DTC was recently published (Park & Yoon 2019).

A summary of the studies analyzing dynamic risk assessment in patients treated with total thyroidectomy or lobectomy without RA can be seen in Table 3.

Dynamic risk assessment in children and adolescents

DTC is the most common endocrine tumor in children and adolescents; it represents 1% of tumors in prepubertal children, a percentage that increases to 7% in patients between 15 and 19 years old (Verburg et al. 2017). In the pediatric age, DTC frequently is diagnosed as an advanced disease that exceeds the limits of the thyroid gland, with loco-regional metastases to the cervical lymph nodes and very frequently, with distant metastases to the lungs. These children also have a high frequency of recurrence, especially in those under 10 years old (Verburg et al. 2017). Despite this fact, the 10-year survival rate is 98–99% (Demidchik et al. 2006).

Until now, there is no widely validated postoperative staging system in pediatric patients (Francis et al. 2015).
The AJCC/TNM system has little utility, given that all patients will be classified as stage I (absence of distant metastasis) or II (presence of distant metastasis) (Amin et al. 2017). The 2015 ATA Guidelines for Children with Thyroid Nodules and Differentiated Thyroid Cancer proposes to categorize children into low, intermediate and high risk in order to identify the risk of persistence of the disease and then design the strategy for postoperative follow-up (Francis et al. 2015). In this classification, the low-risk group includes patients with intrathyroidal tumors without lymph node metastases or patients with incidental N1a disease (microscopic metastasis to a ‘small number’ of central neck lymph nodes). The intermediate group is defined as having ‘extensive’ lymph node metastasis in the central compartment or ‘minimal’ number of lymph nodes in the lateral compartment. On the other hand, patients with locally invasive tumors, ‘extensive’ lateral lymph node involvement, or distant metastases belong to the high-risk group. This classification is unclear, given that the number of lymph nodes considered as ‘minimal’ or ‘extensive’ involvement is not specified. In a cohort of 77 DTC patients <18 years, 64 with total thyroidectomy and RA and 13 with lobectomy, the dynamic risk stratification system was assessed using the classification of RR for pediatric patients proposed by the ATA guidelines (Sung et al. 2017) and the responses to treatment proposed by the ATA guidelines for adult patients (Haugen et al. 2016). The authors considered minimal N1a disease when metastatic foci was <0.2 cm and involved in <5 lymph nodes as low risk and minimal N1b disease with <10 lymph nodes with metastatic foci <3 cm as intermediate RR. The initial excellent response ranged from 82% for low RR patients, 62% for intermediate-risk patients and 4% for high-risk patients. On the other hand, although the risk of having structural incomplete response at the end of follow-up was significantly higher in the high-risk group (65%), this difference was not evident between low and intermediate RR patients (12 and 19%, respectively). It seems that the probability of having a structural incomplete response in low RR patients is higher in pediatric patients than in adults, who have a risk of structural disease between 2 and 6% (Sung et al. 2017). However, only 22% of patients in this study had a low RR. Considering the dynamic risk assessment, the excellent response at the end of follow-up was reached in 97% of the patients with an initial excellent response, in 60% of those with an initial indeterminate response, in the 40% of those who had an initial biochemical incomplete response and only 4% of those who showed an initial structural incomplete response. The majority (88%) of patients with a structural incomplete response at the end of follow-up were patients with the same response at initial evaluation. Only 1 patient (3%) with an excellent initial response had recurrence of the disease confirming the impact of the initial response, even in the pediatric population of patients with DTC (Sung et al. 2017).

Similar results were found by subsequent studies in pediatric patients with DTC (Sohn et al. 2017, Kim et al. 2018, Zanella et al. 2018, Remiker et al. 2019). The percentages of structural incomplete responses at the end of follow-up and excellent responses according to the initial assessments were similar to those shown in patients >18 years of age, which suggests that the dynamic risk stratification can also be used in pediatric population. It must be considered that in these cohorts, the majority (56–73%) of patients had >15 years old, a subgroup of patients that would have a more favorable prognosis in comparison with younger patients.

### Table 3 Summary of studies analyzing dynamic risk assessment in differentiated thyroid cancer patients treated without radioiodine ablation.

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of patients</th>
<th>Patients included</th>
<th>Median follow-up (months)</th>
<th>Persistent/recurrent disease based on each response-to-therapy category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Momesso et al. 2016</td>
<td>TT: 320</td>
<td>Low risk: 433</td>
<td>100.5</td>
<td>Exc 0% IR 1.3% BIR 31.6% SIR 100%</td>
</tr>
<tr>
<td></td>
<td>L: 187</td>
<td>Intermediate risk: 74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Park et al. 2017</td>
<td>TT: 64</td>
<td>Low risk: 187</td>
<td>103.2</td>
<td>Exc 0.78 IR1.5% BIR 16.7% SIR 100%</td>
</tr>
<tr>
<td></td>
<td>L: 293</td>
<td>Intermediate risk: 170</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abelleira et al. 2017</td>
<td>TT: 88</td>
<td>Low risk: 78</td>
<td>28</td>
<td>Exc 0% IR 5.6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Intermediate risk: 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee et al. 2018</td>
<td>TT: 26</td>
<td>Low and intermediate risk</td>
<td>124</td>
<td>Exc 7.4% IR 0% BIR 0% SIR 100%</td>
</tr>
<tr>
<td></td>
<td>L: 69</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cho et al. 2018</td>
<td>L: 619</td>
<td>Low risk: 340</td>
<td>103</td>
<td>Exc 1.6% IR 3.8% BIR 2.9% SIR 100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Intermediate risk: 279</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BIR, initial biochemical incomplete response; Exc, initial excellent response; IR, initial indeterminate response; L, lobectomy; SIR, initial structural incomplete response; TT, total thyroidectomy.
Future perspectives

Contributions of molecular markers to the dynamic risk assessment

In recent years, some studies have suggested the use of molecular markers to predict the RR in patients with DTC. Although not routinely recommended for initial postoperative staging, the BRAF mutation and others, such as TERT mutations, would have the potential to refine the RR estimation when interpreted in the context of other clinical-pathological risk factors (Haugen et al. 2016).

In a meta-analysis of 14 publications that included 2470 patients with PTC from nine different countries, the BRAFV600E mutation was associated with higher RR than wild type tumors (24.9 vs 12.6%, P<0.00001 (IC 95% 1.61–2.32) (Tufano et al. 2012). Given that this mutation is usually associated with other clinical-pathological variables such as aggressive histology, lymph node metastases and extrathyroidal extension, it is difficult to determine in which proportion the BRAF mutation is independently associated to an increase in the RR. On the contrary, when the mutational state of BRAF is associated with other clinical factors, it could be useful (Haugen et al. 2016). This is evident in multifocal papillary microcarcinomas with minimal extrathyroidal extension, where the presence of BRAF mutation increases the RR to 20% (Niemeier et al. 2012) (currently categorized as intermediate risk by the 2015 ATA guidelines for adult patients). In a recent study of 620 DTC patients, BRAF status taken together with stimulated Tg levels, measured at the time of RA, improved the performance of the ATA risk stratification system. In particular, considering the low-risk patients group, only positive BRAF V600E mutation combined with a stimulated Tg level measured at the time of RA >8.9 ng/mL was associated with persistent disease (Repaci et al. 2019).

TERT mutation, on the other hand, also turned out to be an independent predictor of DFS and mortality in patients with DTC (Melo et al. 2014). The inclusion of the TERT mutational status in the dynamic risk stratification in patients with DTC increases the prediction of structural disease (Kim et al. 2017b). The authors divided the cohort (n=357) into four groups: (1) patients with excellent response without TERT mutated, (2) patients with indeterminate response without TERT mutated, (3) patients with structural/biochemical incomplete response without TERT mutated or patients with excellent/indeterminate response with TERT mutated and (4) patients with structural/biochemical incomplete response and TERT mutated. The 10-year recurrence-free survival was 94.9, 82.7, 50.2 and 23.1%, respectively (P<0.001), and cancer-specific survival rates were 100, 98.7, 94.2 and 76.9%, respectively. It is interesting to note that patients harboring TERT mutation who achieve an excellent initial response still persist with a high probability of recurrence, although the survival is not less than in the groups with non-mutated TERT (Kim et al. 2017a). This could clarify the discrepancy between some groups that report a 14% structural incomplete response in high RR patients who achieve an excellent response (Tuttle et al. 2010), compared to those who showed that the initial excellent response is independent of the RR (Castagna et al. 2011, Jeon et al. 2018). Following this line of thought, patients with an initial excellent/indeterminate response and TERT mutated (considered as harboring a high RR) should probably have a more strict follow-up. Another interesting point from this study (Kim et al. 2017a) is that patients with TERT mutated who reached a biochemical incomplete or a structural incomplete response have a very poor prognosis. The authors suggest that since this type of mutation has been associated with dedifferentiated tumors and low radioiodine avidity (Kim et al. 2016), patients with TERT mutations who do not present an excellent response might already be considered as potentially radioiodine-refractory and therefore candidates to other systemic therapies when structural disease non-amenable for surgical treatment is found (tyrosine kinase inhibitors or redifferentiation therapy) (Kim et al. 2016). On the other hand, the combination of TERT and BRAF mutations in the same tumor was associated with higher risk of structural incomplete response than when they were analyzed separately (Xing et al. 2014), suggesting that the combination of these mutations is probably more useful for improving the RR prediction.

Interestingly, a study using next-generation sequencing (NGS) found that TERT mutations are subclonal in the few PTCs that harbor them (9%), whereas they are clonal in poorly DTCs and anaplastic thyroid cancers (where the frequency of TERT mutation is 40 and 73%, respectively), denoting selection during tumor evolution (Landa et al. 2016). Simultaneous analysis of cancer-associated genes of interest (BRAF, KRAS, NRAS, PIK3CA, TP53 and so forth) by using NGS could further enhance the understanding of the clinical behavior of DTC and allow a better prediction of the patient outcomes (Landa et al. 2016, Bandoh et al. 2018).
Circulating miRNAs are emerging as new blood markers of poorer prognosis in DTC. Gómez-Pérez et al. determined the levels of miR-146, miR-221 and miR-222 in the peripheral blood of 60 DTC patients in correlation with the dynamic risk stratification. The authors found that elevated miR-222 and miR-146 levels were associated with poorer outcomes of the disease, suggesting that these miRNA may have a prognostic value in the management and follow-up of DTC patients (Gómez-Pérez et al. 2019).

Existing and ongoing prospective trials

Recently, the data from the medium-term follow-up from the ESTIMABL1 trial were published (Schlumberger et al. 2018). This was a prospective randomized, phase 3 study in patients with low-risk thyroid cancer that compared four strategies of postoperative RA in a 2 × 2 factorial design: half of the patients were stimulated with recombinant human thyroid-stimulating hormone (rhTSH), half with 4 weeks withdrawal of thyroid hormone; within each group, half of the patients were given a low-activity (1.1 GBq; 30 mCi) or high-activity (3.7 GBq; 100 mCi) radioiodine. At 6–8 months, thyroid ablation was complete in 631 (92%) of 684 patients, with equivalent proportions for all four ablation strategies (Schlumberger et al. 2012). The outcome at 5 years showed only 2% of structural incomplete response, confirming the retrospective data already published. Additionally, this study showed that the initial disease status of each patient was strongly correlated with the stimulated thyroglobulin concentration at ablation, only 2% of patients with stimulated Tg levels lower than 5 ng/mL at the moment of RA had a structural incomplete response during follow-up (Schlumberger et al. 2018). These data are also consistent with the HiLo study, which showed a larger frequency of SIR (5%) after a median follow-up of 6.5 years. This is mainly attributed to the fact that there were a proportion of included patients with an intermediate RR (Dehbi et al. 2019). Interestingly, recurrences were still seen in HiLo study beyond 5 years suggesting that a simplified surveillance protocol with the use of serum Tg alone might be considered after 5 years of follow-up for some or even all patients (Dehbi et al. 2019).

These two recent prospective trials in combination with previous observational studies showed definitive evidence that low-dose RA is an acceptable therapy for DTC patients. Two large prospective trials are ongoing, the IoN trial (NCT01398085) and ESTIMABL2 (NCT01837745), including low and low-to-intermediate-risk patients (IoN also includes patients with minimal extrathyroidal extension and N1a disease) randomly assigning to have either low-dose (30 mCi) RA or no RA, with DFS as a major endpoint. These studies will be key to clarify the need for RA in most patients with DTC.

Conclusions

The dynamic risk stratification constitutes a paradigm shift in the management of patients with DTC. In recent years, we moved from having a single treatment for all patients (total thyroidectomy, RA, regular diagnostic whole body scans, etc.) to an individualized treatment initially supported by the RR, as a result of the analysis of the first data obtained after the surgery, the histopathological study and the clinical and perioperative serum Tg/anti-Tg levels, a situation that will be modified based on the responses obtained throughout the follow-up. This approach avoids excessive treatment and intensive follow-up in the vast majority of patients who will have a very good prognosis (low risk and most of those with intermediate risk of recurrence) and, on the other hand, focus therapeutic efforts on the few patients with worse prognosis (high RR patients).

The validation of this dynamic stratification system in patients with lobectomy and patients who did not receive RA is an incentive for many endocrinologists to opt for this practice in patients with low and even for those with an intermediate RR. On the other hand, adolescents and children over 10 years of age with DTC may also be candidates for this approach, although the studies that support this approach are scarce.

In the future, more studies showing the impact of the molecular biology analysis of the surgical piece will probably further refine the RR prediction. Ongoing prospective studies will provide a more solid basis for therapeutic decisions regarding dynamic risk in this complex management of the individual patient with DTC.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of this review.

Funding

This work did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.
References


Castagna MG, Maino F, Cipri C, Belardini V, Theodoropoulou A, Cevenini G & Pacini F 2011 Delayed risk stratification, to include the response to initial treatment (surgery and radioiodine ablation), has better outcome predictivity in differentiated thyroid cancer patients. European Journal of Endocrinology 165 441–446. (https://doi.org/10.1530/EJ-11-0466)


La Greca A, Pitoia F & Tuttle RM 2017 The ‘broken chair’ in patients with differentiated thyroid cancer. Endocrine 57 359–360. (https://doi.org/10.1007/s12020-017-1345-y)


Postoperative thyroid-stimulating hormone levels did not affect recurrence after thyroid lobectomy in patients with papillary thyroid cancer. Endocrinology and Metabolism 34 150–157. (https://doi.org/10.3803/EnM.2019.34.2.150)


Park JH & Yoon JH 2019 Lobectomy in patients with differentiated thyroid cancer: indications and follow-up. Endocrine-Related Cancer 26 R381–R393. (https://doi.org/10.1530/ERC-19-00885)


Pitoia F, Bueno F, Urciuoli C, Abelleira E, Cross G & Tuttle RM 2013 Outcomes of patients with differentiated thyroid cancer risk-stratified according to the American Thyroid Association and Latin American Thyroid Society risk of recurrence classification systems. Thyroid 23 1401–1407. (https://doi.org/10.1089/thy.2013.0011)


Pitoia F, Jerkovich F, Smulever A, Brenta G, Bueno F & Cross G 2017 Should age at diagnosis be included as an additional variable in the risk of recurrence classification system in patients with differentiated thyroid cancer. European Thyroid Journal 6 160–166. (https://doi.org/10.1159/000435445)


Tarasco HD & Tuttle RM 2016 A risk-adapted approach to follow-up in differentiated thyroid cancer. Rambam Maimonides Medical Journal 7 e0004. (https://doi.org/10.5041/RMMJ.10231)


differentiated thyroid cancer after total thyroidectomy and radioactive iodine remnant ablation: using response to therapy variables to modify the initial risk estimates predicted by the new American Thyroid Association staging system. *Thyroid* **20** 1341–1349. (https://doi.org/10.1089/thy.2010.0178)


Vaisman F, Shaha A, Fish S & Tuttle MR 2011a Initial therapy with either thyroid lobectomy or total thyroidectomy without radioactive iodine remnant ablation is associated with very low rates of structural disease recurrence in properly selected patients with differentiated thyroid cancer. *Clinical Endocrinology* **75** 112–119.

Vaisman F, Tala H, Grewal R & Tuttle RM 2011b In differentiated thyroid cancer, an incomplete structural response to therapy is associated with significantly worse clinical outcomes than only an incomplete thyroglobulin response. *Thyroid* **21** 1317–1322. (https://doi.org/10.1089/thy.2011.0232)


Received in final form 20 July 2019
Accepted 7 August 2019
Accepted Preprint published online 7 August 2019